Assessing the performance of density functional theory in optimising molecular structure parameters

Assessing the performance of density functional theory in optimizing molecular crystal structure parameters

Jack Binns, Mary R. Healy, Simon Parsons* and Carole A. Morrison*

The School of Chemistry, University of Edinburgh, The King's Buildings, West Mains Road, Edinburgh, EH9 3JJ, UK.

Abstract

This paper assesses the performance of plane-wave density functional theory calculations at returning reliable structural information for molecular crystal structures where the primary intermolecular interactions are either hydrogen bonding or dispersion interactions. The computed structures are compared with input structures obtained from the Cambridge Structural database, and assessed in terms of crystal packing similarities, unit cell volume and shape, short contact distances and hydrogen bond distances. The results demonstrate that the PBE functional with the TS dispersion correction is capable of returning reliable full structural optimisations, in which both atomic positions and unit cell vectors are free to optimise simultaneously.

1. Introduction

The last few years has seen a rapid rise in the popularity of first-principles condensed matter simulations in structural and materials chemistry applications. (Burke, 2012) There are many reasons for this, with a particularly appealing one being the prospect of predicting the crystal structures and properties of compounds from the molecular geometry alone, for which the accurate simulation of lattice energies is an all-important step. (Neumann, 2005; Asmadi, 2009, Pickard, 2011). Condensed matter simulations also have a fundamental role to play in the crystallographic refinement process, for it can provide information relating to parameters which would otherwise have to be fixed at some assumed values. Examples of this include the refinement of limited data sets collected under conditions of high temperature and/or pressure, (Funnell, 2010) and in the treatment of crystal structure disorder and diffuse scattering. (Morrison, 2002; Thomas 2011) And in materials chemistry applications, a calculation following on from a crystal structure determination can yield valuable physical properties, such as lattice energies, intermolecular interaction energies, (Morrison, 2003; Hunter 2013) and other electronic structure phenomenon such as band structure plots. Simulating NMR and phonon spectra can now also be considered routine calculations (Yates 2007, Reilly, 2013a). The rise in popularity of first-principles condensed matter calculations must also, of course, be attributed to the improvements in software and hardware that have rendered complex calculations accessible by non-specialists in realistic timescales. Thus crystallographers now have at their disposal a technique which in many regards is complementary to their experimental methods.

For first-principles condensed matter calculations the most tractable solution at present is density functional theory (DFT), as the alternatives based on Hartree-Fock theory are significantly more computationally expensive, especially when coupled to a plane-wave basis set. In DFT the total ground-state energy of a system can be found from the electron density, ρ, which can be written using the following expression

$$E_{tot}[\rho] = T_k[\rho] + V_{ne}[\rho] + V_{nn}[\rho] + V_{ee}[\rho] + E_{XC}[\rho]. \quad (1)$$

The terms on the right-hand side of the equation are respectively, (i) the kinetic energy for a fictitious system of *N* non-interacting electrons, the potential due to the interaction between (ii) the electrons and the nuclei, (iii) pairs of nuclei and (iv) pairs of electrons. The last term is the exchange/correlation functional, which describes the components of the earlier terms that cannot currently be treated exactly. Knowledge of what form this functional should take is only known exactly for a free electron gas. For all other 'real' systems approximations must therefore be made. The functionals that are commonly applied in molecular simulation are based on the generalised gradient approximation (GGA), with two of the most commonly used examples being PBE and PW91. (Perdew, 1996; Perdew 1992)

The task facing first-principles condensed matter simulations is a challenging one. To be useful it has to accurately describe the whole range of atomic interaction potentials, spanning from ionic bonding (typically many hundreds of kJ/mol) through covalent interactions (typically a few hundreds of kJ/mol) to the weak intermolecular interactions including hydrogen-bonding (less than *ca.* 50 kJ/mol). One of the intrinsic failings of DFT is in the description of dispersion forces (London or attractive van der Waals forces) which arise from electron correlation at long range. These are often very weak interactions (typically only a few kJ/mol), but in a molecular crystal structure there can be lots of them, so simply omitting them from a simulation runs the risk of incurring substantial errors. GGA functionals like PBE and PW91 utilise only the local density and its gradient, and are thus unable to describe these attractions properly. Since it is known that the leading dispersion term is the instantaneous dipole-dipole interaction, which can be described as $1/r^6$ (where *r* is the internuclear separation between a pair of atoms A and B), a convenient remedy is to bolt-on an additional term to the DFT energy, such that:

$$E_{tot} = E_{DFT} + E_{disp.} \quad (2)$$

The dispersion interaction energy is given by:

$$E_{disp} = \sum_{\mathrm{A,B}} \frac{f(r_{\mathrm{AB}}, \mathrm{A}, \mathrm{B}) C_6^{\mathrm{AB}}}{r_{\mathrm{AB}}^6} \quad (3)$$

where the dipole-dipole dispersion coefficents terms C_6^{AB} depend on the atoms A and B. (Grimme, 2011) Thus the total dispersion energy can now be calculated by summing over all pairs of atoms in the model. The choice of parameterisation for C_6^{AB} is critical, and a number of different approximations have emerged. At the simplest level, values for C_6^{AB} can be obtained from experimental data (such as ionisation potentials and atom polarisabilities) and are held at fixed values throughout the simulation. Higher levels attempt to obtain more accurate values by including the effects of the crystallographic chemical environment. The function $f(r_{\mathrm{AB}}, \mathrm{A}, \mathrm{B})$ is a damping function, which acts to switch off the dispersion correction at short range to allow bonding distances to be solely described by DFT. The resulting schemes are often referred to as DFT-D. An excellent review of this field is available (Klimeš, 2012). Specific correction schemes include Ortmann, Bechstedt and Schmidt's 'OBS' (Ortmann, 2006), Grimme's 'G06' (Grimme, 2006) and 'DFT-D3' (Grimme, 2010), and Tkatchenko and Scheffler's 'TS' (Tkatchenko, 2009) schemes. Of these, the OBS scheme is the most basic, with the data required to construct the C_6^{AB} terms coming from experiment and the damping function constructed in such a way as to reproduce the *c*-lattice vector of graphite. The results from this model are likely to be less transferrable to other systems. The G06 scheme has fixed (pre-defined) values for the C_6^{AB} coefficients, originating from ionization potentials and static polarizabilities for isolated

atoms, whereas the values used in the DFT-D3 and TS schemes are system specific, having been computed from first principles (in the latter from a Hirshfield partitioning of the computed electron density of the molecule under study). The G06, DFT-D3 and TS correction schemes use similar damping functions.

It is important to be aware, however, that this pairwise approach only considers dispersion interactions in a crystal lattice at a local (short range) level. The effects of longer range dipole-dipole fluctuations are not considered. However, the recently developed 'many-body dispersion' (MBD) method does take this into account by extending the TS method to include long-range screening effects in the calculated effective atomic polarizabilities (Schatschneider, 2013; Tkatchenko 2012; DiStasio, 2012). It has been shown that this approach can half the mean absolute error observed when calculating lattice energies, compared to results obtained where just the pairwise TS correction is applied (Reilly, 2013a). It also has a demonstrated effectiveness in correctly ordering the relative thermodynamic stabilities of polymorphs of extended molecular systems (DiStasio, 2012; Marom, 2013).

Given the range of applications that first-principles condensed matter simulations now enjoy in the field of crystallography, it is perhaps surprising that there are not more papers in the literature documenting a systematic study on the ability of DFT and the various dispersion correction schemes to optimize crystal structure parameters. Some examples do exist, and these include a polymorphism studies in para-diiodobenzene (Pedone, 2012), oxyalyl dihydrazide (Presti, 2014), and a study on a range of extended solids that included noble gas solids, molecular crystals, layered materials and covalent solids (Al-Saidi, 2012). It is more often the case that papers in this area have as their primary focus the ability of the simulation to reproduce lattice energies, bulk moduli, and other thermodynamic quantities, with less emphasis placed on structural data. (Otero-de-la-Roza, 2012; Zheng, 2012; Reilly 2013a; Reilly 2013b) The reason for this is simply that those properties offer much more sensitive indicators of the success (or failure) of a particular dispersion correction scheme, than the structural parameters do.

DFT method development is, and is likely to remain, an active area of research. So while it would be unwise to attempt to construct a final statement on the performance of DFT functionals, it is however timely to present a summary of their current performance, with a primary focus being to quantify their ability to model crystal structure parameters. To this end, we have constructed two datasets of organic molecular crystal structures obtained from the Cambridge Structural database (CSD). Class I contains structures where the packing forces are dominated by hydrogen bonding, while Class II contains structures in which dispersion is the main source of intermolecular interaction. We have undertaken a series of geometry optimisation calculations using the DFT functionals and dispersion corrections schemes as currently available in the condensed matter simulation package CASTEP. They are PBE, (Perdue, 1996), PBEsol (Perdue, 2008) (a revision of the original PBE functional specifically designed for densely-packed solids) and PW91, (Perdue 1992) along with three different dispersion correction schemes, OBS (Ortman, 2006), G06 (Grimme, 2006) and TS (Tkatchenko, 2009). For a subset of the Class II structures we have also applied the MBD scheme. Note we have not attempted to quantify the behaviour of other functionals, such as hybrid, range-separated or meta-GGA, which could in principle offer improved performance over standard GGA functionals.

The results we present from these simulations will be discussed in the following order. First, we report on the crystal packing similarity (CPS) scores (Chisholm, 2005), which represent an RMS fit obtained in overlaying 15 molecules from the simulated and experimental crystal lattices. This provides a quick visual handle on how the functionals and dispersion correction schemes compare. We then follow this up with an analysis of how the simulations compare when attempting to optimise the unit cell vectors (unit cell

volumes and cell shapes). Finally, we offer a statistical analysis of how the simulations reproduced bond distances, and comment on reliability of simulating hydrogen bond distance parameters.

2. Computational Method

Structure selection. The molecular crystal structures were selected from the CSD using CONQUEST 1.13. (Allen, 2002; Bruno, 2002) Only crystal structures of organic molecules that were derived from single-crystal data collections below 50 K (to minimise the effects of molecular vibrations on the resulting structures), that were free from disorder, that had unit cell volumes below 1000 $Å^3$ and that reported *R*-factors below 0.08 were considered. For the Class I set, structures also had to contain hydrogen bonding interactions, defined as X-H…Y, with X…Y $\leq$ 2.5 Å where X and Y = O or N; for the Class II set this interaction must be absent. The Class I set comprises a total of 17 structures, all of which originate from neutron diffraction experiments; the Class II set comprises 13 structures, of which 9 are derived from neutron diffraction (the remainder from X-ray diffraction). The compound names are given in Figure 1; the corresponding CSD codes are listed in the Supplementary Information.

Computational procedure. All calculations were carried out using the plane-wave DFT computational package CASTEP 5.501, (Clark, 2005; McNellis, 2009) and CASTEP 6.11 (for the MBD correction scheme simulations (Tkatchenko, 2012)), with input files prepared using Materials Studio 5.5 and 'on-the-fly' pseudopotentials generated directly by the CASTEP package. For each molecular crystal, the first calculation step involved the determination of a suitable energy cut-off for the plane-wave basis set. This was carried out by repeated single-point energy calculations with an increasing energy cut-off until convergence of $dE_{tot}/dE_{cut} \leq 0.003$ eV per atom was achieved. For all simulations sampling of the Brillouin zone was achieved using a Monkhorst-Pack grid, (Monkhorst, 1976) with parameters chosen to result in a separation between *k*-points of generally no more than 0.08 $Å^{-1}$. Standard convergence criteria were applied (maximum change in system energy 2.0×10^{-5} eV/atom, maximum force 0.05eV/Å, maximum root-mean square (RMS) atomic displacement 0.002 Å, and maximum stress 0.1 GPa). Full details of the computational models chosen, along with sample convergence criteria plots, can be found in the Supplementary Information.

Two different geometry optimisation procedures, both of which took the experimental structure as input and applied the appropriate space group symmetry constraints, then followed. In the first, only the atomic positions were allowed to optimize, while unit cell parameters were fixed at experimental values. In the second, atomic positions and unit cell vectors were allowed to vary simultaneously. This was done in order to reflect the contrasting approaches currently taken in the simulation of crystalline materials. One application of DFT calculations is to provide data suitable for use in crystal structure refinements where the intensity data are incomplete (as in high-pressure experiments), suffer from peak overlap (as in powder diffraction data), or are compromised in some other way. In such cases there is usually no need to optimise unit cell parameters since these are known accurately from the experimental data. What is required is a source of geometric restraints or constraints, and fixed-cell DFT optimisations provide this. Fully DFT-optimised structures in which both cell parameters and coordinates are allowed to vary are needed in cases where no information of the unit cell is available from experimental sources, for example in a completely *ab initio* study of the effect of pressure or temperature on a crystal structure.(Chen, 2013) The failure to account for dispersion interactions in the past often resulted in unrealistic expansion of the unit cell vectors, and our aim in performing the cell+coordinate optimisations was to ascertain whether the addition of the classical dispersion correction schemes now offers a reliable route for obtaining an accurate molecular crystal structures by simulation alone.

All structures were optimised using the DFT functionals PBE (Perdue, 1996), PW91 (Perdue, 1992) and PBEsol. (Perdue, 2008) We tested the dispersion correction schemes available in CASTEP 5.501 and 6.11, which are G06, (Grimme, 2006) TS, (Tkatchenko, 2009) for use alongside PBE, and OBS (Ortmann, 2006) which is available for PW91. We have also applied MBD to a subset of Class II (dispersion dominated) structures.

Data analysis. Output crystallographic information files '.cif' were generated automatically by the CASTEP code and used for analysis in CCDC Mercury 3.0 (Macrae, 2008; Macrae 2006) to give the following parameters: bond lengths, O…H contact lengths, N…H contact lengths, unit cell volumes and crystal packing similarities (CPS). This latter quantity measures the root mean square deviation between a reference experimental structure and an optimised structure by superimposing the atomic coordinates for a cluster of 15 molecules derived from each. (Chisholm, 2005; Macrae, 2008; Macrae 2006) A value of zero corresponds to a perfect correlation. Note the positions of the hydrogen atoms were omitted from this analysis.

3. Results and Discussion

3.1 Crystal packing similarities (CPS)

An overall comparison of the general performance of functionals and dispersion correction schemes is available from the crystal packing similarities (CPS) in Figure 1 and in Table 1, with full numerical data reported in Tables S3 and S4 in the Supplementary Material. As expected, CPS deviations are substantially larger for those simulations where the unit cell parameters were allowed to optimize.

For the fixed-cell optimisations the best overall performance was obtained with the PBE functional. Inclusion of the TS dispersion term appears to have a negligible effect on both classes of compound, but the G06 scheme appears to degrade performance slightly for Class I. This demonstrates that, on the whole, the damping functions employed in the dispersion correction terms are working appropriately, with all short-range geometrical interactions resulting from the underlying pure DFT functional alone. Thus, from a purely geometrical perspective, fixed-cell optimisations using functionals without dispersion corrections are just as reliable as fixed-cell optimisations that make use of the correction.

By contrast, use of a dispersion correction term is extremely important in obtaining reliable optimisation results where both the structure and the unit cell dimensions are allowed to vary. The dispersion-corrected schemes all offer substantial (and in some cases very dramatic) improvement over the pure DFT functionals. Pure DFT functionals perform equally badly in variable-cell optimisations for both Class I and II, while the addition of the dispersion correction scheme renders structures that are dominated by dispersion interactions as no less accurate than those that are dominated by hydrogen bonding interactions. Though they are frequently neglected in descriptions of hydrogen-bonded crystal structures, dispersion interactions are always present, and the similar performance of the functionals with and without dispersion corrections in both compound groups is a reflection of the importance of dispersion in all molecular crystal structures.

The outcome of this work is that PBE-TS appears to offer the strongest performance across the board (with an overall CPS RMS fit to within 0.02 and 0.08 Å for fixed and optimized unit cell simulations, respectively), with PBE-G06 a close second. Note the poor performance of PW91-OBS must be largely due to the OBS correction, in the same way that the superior performance of PBE-TS is largely due to the TS correction, as the underlying data from the pure PW91 and PBE optimization results are not so wholly different. PBE-MBD simulations on three dispersion dominated structures (see Table S4 in the Supplementary material) return further minor improvements on PBE-TS. Such deviations that –remain are, of course, due in part to deficiencies in the functionals and dispersion corrections, but also to the effects of thermal vibrations in the

experimental data. We have tried to minimise the latter by selecting crystal structures determined at low temperature ($T \leq 50$ K), but there is still a possibility that thermal motion could be responsible for some of the differences between the calculated (0 K) unit cell parameters and the thermally-averaged experimental structures.

3.2 Unit cell volumes and shapes

Plots of experimental versus unit cell volume obtained from the variable cell optimisations are given in Figure 2. Pure DFT functionals overestimate the optimised cell volumes, in some cases by as much as 50%, with PW91 generally fairing the worst. PBEsol offers some improvement, but the values obtained by inclusion of a dispersion correction most consistently reproduce experimental values. Both TS and G06 reproduce experimental values accurately, with data points for TS generally lying slightly closer to the $y=x$ line [Class I: PBE-TS gradient of best fit line m = 1.001(5), PBE-G06 m = 0.966(6); Class II: PBE-TS m = 1.012(16), PBE-G06 m = 0.963(15)]. From these results we also observe that the calculated unit cell volumes for the Class I compounds are more accurately modelled than those of Class II. For both compound classes PW91-OBS tends to overestimate the strength of the intermolecular interactions, consistently reporting an optimised unit cell volume that is too small. PBE-MBD does not offer significantly improved calculated unit cell volumes over the results obtained for PBE-TS for the three Class II structures sampled (see Table S6 in the Supplementary Material).

In judging the performance of the different calculations, simple consideration of unit cell volumes could miss a case where a fortuitously accurate calculated volume arose because one unit cell dimension was too small, while another was too large. Comparison of observed and calculated values of the longest unit cell vector divided by the shortest, shown in Figure 3, should reveal this situation. Consistent with the results discussed above for the cell volumes, the data in Figure 3 show that the DFT-D functionals perform consistently better than the pure DFT functionals for both classes of compound. The TS correction scheme appears to score slightly better, and that, again, the Class I crystal structures are more accurately modelled than those of Class II [Class I: PBE-TS m = 1.001(3), PBE-G06 m = 1.003(4); Class II: PBE-TS m = 1.010(13), PBE-G06 m = 1.018(12)]. As with unit cell volume determination, PBE-MBD does not appear to offer significant improvement over PBE-TS (see Table S8 in the Supplementary Material).

3.3 Calculated bond lengths

Table 2 lists the gradient of the best-fit line, m, and associated standard deviation for the nine most prevalent bond types (namely, rC-O, rC-C, aromatic rC-C, rN-N, rC-N, rC-H, rN-H and rO-H) collated into two test groups, depending on whether or not the bond involves hydrogen. This comparison shows that all functionals are capable of returning optimised values for the internal bond lengths to very similar levels of accuracy, and that accuracy is not affected by allowing the unit cell vectors to vary. This was as expected, as it indicates that the damping functions applied in the dispersion correction schemes are performing properly, with all short-range interactions modelled purely by DFT. Note the simulations appear to consistently return values for the hydrogen-containing bond distances that are significantly longer than the experimental values. This is despite the fact that the experimental structures are obtained almost entirely from neutron diffraction experiments.

Breaking the data set down into individual bond types reveals that the choice of functional has very little bearing on the optimised values obtained. All bond types without hydrogen have maximum deviations that lie comfortably within the standard deviations for the experimental distributions. However, the bond types

with hydrogen display maximum deviations that are greater than the standard deviations for the experimental dataset for all functionals under both fixed and free cell conditions, with PBEsol fairing the worst. While these maximum deviations are clear outliers the overall trend points to a systematic lengthening of all hydrogen-containing bonds. This may simply be a consequence of ignoring librational motion in the analysis of the experimental data which results in measured bond lengths that are too short, particularly for those involving hydrogen. In order to quantify this we have used PLATON (Spek, 2009) to perform thermal libration corrections for the C-H bonds present in the crystal structure of perdeuteromalonic acid (a member of the Class I set). Corrections of the order of around +0.01 Å were obtained, which does account for most of the discrepancy noted for C-H bond distances recorded in Table 2.

3.4 Calculated *r*O…H and *r*N…H interactions

We now turn our attention to the calculated O…H and N…H hydrogen bond interactions present in our Class I compounds, and report the gradients of straight-line fits to the experimental data in Table 4. We divide our data into deuterated and non-deuterated sets, as the simulations consistently return more accurate values when compared against deuterated compounds.

In the optimised-cell calculations the pure functionals give O…H/D contact lengths that are too long, while all corrected functionals suffer from overbinding to some extent, leading to linear fit gradients less than 1. The best performance is again exhibited by PBE-TS, with PBE-G6 a close second. In the fixed-cell calculations all methods yield distances which are too short, with addition of a dispersion correction for PBE making little difference.

Trends for the *r*N...H/D contacts are similar, although in this case the best results amongst the variable-cell calculations are obtained for the PBE and PW91 without dispersion corrections. This rather anomalous result can probably be attributed to the size of the data set (only 12 observations); this point aside PBE-TS again appears to report the most faithful simulation of the experimental structures.

4.0 Conclusions

In this work we have investigated the ability of various DFT and DFT-D functionals to simulate the geometric parameters of crystal structures that are either dominated by hydrogen bonding or dispersion interactions. For atom-only minimisations the inclusion of a dispersion correction to the DFT functional offers no improvement on the resulting structure, as all short-range contacts are computed by the pure DFT functional alone and the long-range intermolecular contacts (where the dispersion correction is really needed) are constrained by fixing the unit cell dimensions at the experimental values. For simulations where the unit cell vectors were free to vary, the inclusion of a dispersion correction scheme is crucial to obtaining accurate results. We found that the TS scheme, when coupled to the PBE functional offers a considerable improvement over non-corrected functionals, giving the lowest CPS RMS values of 0.08(6) Å and 0.06(6) Å for Class I and Class II, respectively. As well as accurate values for unit cell volumes being obtained, the shapes of unit cells can now also be reliably optimised, with PBE-TS offering a slight lead over PBE-G06. From a rather limited sample, it appears that PBE-MBD, while very important for simulating lattice energies to within chemical accuracy, does not offer further substantial improvement over geometries returned by PBETS. All functionals performed equally well when it comes to obtaining optimised values for short-contact distances such as *r*C=O, *r*C-C etc, with the exception of PBEsol, which consistently returned distances that were too short. The comparison to short contacts that contain hydrogen is less good, with theory always predicting bond lengths that are too long, but this may be due to the omission of thermal librational corrections in the treatment of experimental diffraction data. In

replicating intermolecular contacts PBE-TS gives distances that are consistently closer to the experimental values in both fixed and optimised cell conditions, with PBE-G06 ranking not far behind. We therefore conclude that full structural optimisations, in which both atomic positions and unit cell vectors are free to optimise simultaneously, are now attainable for the types of compounds included in this study.

One final comment worth noting here is that the addition of the dispersion correction scheme resulted in markedly shorter compute times. The reason for this is simply that the pure DFT functionals all report volumes that are considerably too large, and therefore require more cycles of optimisation to reach convergence, with each cycle rendering the simulated structure a poorer match with the experimental input geometry. This is therefore a rather rare example where a more complex computational modelling approach actually gives rise to quicker simulations.

References

Al-Saidi, W. A., Voora, V. K. & Jordan, K. D. (2012), *J. Chem. Theory Comput.,* **8**, 1503-1513.

Allen, F. H. (2002). *Acta. Cryst*, **B58**, 380-388.

Asmadi, A., Neumann, M. A., Kendrick, J., Girard, P., Perrin, M. A. & Leusen, F. J. J. (2009). *J. Phys. Chem. B.*, **113**, 16303-16313.

Bruno, I. J., Cole, J. C., Edington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). *Acta. Cryst.*, **B58**, 389-397.

Burke, K. (2012). *J. Chem. Phys*, **136**, 150901-150910.

Chen, L., Mowat, J. P. S., Fairen-Jimenez, D., Morrison, C. A., Wright, P. A. & Duren, T. (2013), *J. Am. Chem. Soc.,* **135**, 15763-15773.

Chisholm J. A. & Motherwell, S. (2005). *J. Appl. Cryst*. **38**, 228-231.

Clark, S. J., Segall, M. D., Pickard, C. J., Hasnip, P. J., Probert, M. J., Refson K. & Payne, M. C. (2005). *Zeitschrift fuer Kristallographie*, **220**, 567-570.

DiStasio Jr. R. A., von Lilienfeld, O. A. & Tkatchenko, A. (2012), *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14791-14795.

Funnell, N. P., Dawson, A., Francis, D., Lennie, A. R., Marshall, W. G., Moggach, S. A., Warren J. E. & Parsons, S. (2010), *Cryst. Eng. Comm.*, **12**, 2573-2583.

Grimme, S., Antony J., Ehrlich, S. and Krieg, H. (2010), *J. Chem. Phys.*, **132**, 154104-154122.

Grimme, S. (2011). *Comput. Mol. Sci.*, **1**, 211-228.

Grimme, S. (2006). *J. Comput. Chem.,* **27**, 1787-1799.

Hunter, S., Sutinen, T., Parker, S. F., Morrison, C. A., Williamson, D. M., Thompson, S., Gould, P. J., & Pulham, C. R. (2013), *J. Phys. Chem. C.,* **117**, 8062-8071.

Klimeš, J. & Michaelides, A. (2012). *J. Chem. Phys.,* **137**, 120901-1209012.

McNellis, E. R., Meyer, J. & Reuter, K. (2009) *Phys. Rev. B*, **80**, 205414-205413.

Macrae, C. F., Bruno, I. J., Chisholm, J. A., Edgington, P. R., McCabe, P., Pidcock, E., Rodriguez-Monge, L., Taylor, R., van de Streek, J. & Wood, P. A. (2008). *J. Appl. Cryst.* **41**, 466-470.

Macrae, C. F., Edgington, P. R., McCabe, P., Pidcock, E., Shields, G. P., Taylor, R., Towler, M. & van De Streek, J. (2006). *J. Appl. Cryst*. **39**, 453-457.
Marom, N., DiStasio Jr., R. A., Atalla, V., Levchenko, S., Reilly, A. M., Chelikowsky, J. R., Leiserowitz, L. and Tkatchenko, A. (2013), *Angew. Chem. Int. Ed.* **52**, 6629-6632.

Monkhorst, H. J. & Pack, J. D., (1976). *Phys. Rev. B*, **13**, 5188-5192.

Morrison, C. A., Wright D. S., & Layfield, R. A. (2002). *J. Am. Chem. Soc*., **124**, 6775-6780.

Morrison, C. A. & Siddick, M. M. (2003). *Chem. Eur. J.*, **9**, 628-634.

Neumann, M. A. & Perrin, M. A. (2005). *J. Phys*. *Chem. B*., **109**, 15531-15541.

Ortmann, F., Bechstedt, F. &Schmidt, W. G. (2006). *Phys. Rev. B*, **73**, 205101-205110.

Otero-de-la-Roza, A. & Johnson, E. R. (2012), *J. Chem. Phys.* **137**, 054103-054112.

Pickard, C. J. & Needs, R. J. (2011), *J. Phys. Condens. Matter,* **23**, 053201-053224.

Pedone, A., Presti, D. & Menziani, M. C. (2012), *Chem. Phys. Lett.,* **541**, 12-15.

Perdew, J. P., Burke K. & Ernzerhof, M. (1996). *Phys. Rev. Lett.* **77**, 3865-3868.

Perdew, J. P. & Wang, Y. (1992). *Phys. Rev. B.,* **45**, 13244-13249.

Perdew, J. P., Ruzsinszky, A., Csonka, G. I., Vydrov, O. A., Scuseria, G. E., Constantin, L. A., Zhou X. & Burke, K. (2008). *Phys. Rev. Lett.* **100**, 136406-136409.

Presti, D., Pedone, A., Menziani, M. C., Civalleri, B. & Maschio, L. (2014) *CrystEngComm,*
DOI: 10.1039/C3CE41758A.

Reilly A. M. & Tkatchenko, A. (2013a). *J. Chem. Phys.,* **139**, 024705-024717.

Reilly A. M. & Tkatchenko, A. (2013b). *J. Phys. Chem. Lett.,* **4**, 1028-1033.

Schatschneider, B., Liang, J.-J., Reilly, A. M., Marom, N., Zhang, G.-X. & Tkatchenko, A. (2013). *Phys. Rev. B,* **87**, 060104-060108.

Spek, A. L. (2009). *Acta Cryst.*, **D65**, 148-155

Thomas, L. H., Craig, G. A., Morrison, C. A., Reilly A. M. & Wilson, C. C. (2011). *Cryst. Growth Des*. **11**, 2045-2049.

Tkatchenko, A. & Scheffler, M. (2009). *Phys. Rev. Lett.,* **102**, 073005-073008.

Tkatchenko, A., DiStasio Jr., R. A., Car, R., Scheffler, R. (2012). *Phys. Rev. Lett.* **108**, 236402-236406.

Yates, J. R., Pickard, C. J. & Mauri, F., (2007), *Phys Rev B.* **76**, 024401-024412.

Zheng, Z., Zhao, J., Sun Y., & Zhang, S. (2012). *Chem. Phys. Lett.,* **550**, 94-98.

Table 1 Average crystal packing similarity (CPS, Å) values obtained for Class I and Class II compounds

	Class I		Class II	
	Fixed unit cell	Optimized unit cell	Fixed unit cell	Optimized unit cell
PBE	0.018(15)	0.46(20)	0.016(14)	0.53(15)
PW91	0.03(4)	0.50(21)	0.014(11)	0.60(22)
PBEsol	0.06(4)	0.21(14)	0.018(17)	0.14(11)
PBE-TS	0.019(13)	0.08(6)	0.016(14)	0.06(6)
PBE-G6	0.03(3)	0.09(4)	0.015(13)	0.14(7)
PW91-OBS	0.08(11)	0.29(16)	0.04(6)	0.38(17)

Table 2 The gradient of the best-fit line, *m*, for the most prevalent intramolecular bond types observed in Class I and Class II compounds obtained for each DFT functional.

	Bonds without hydrogen	
Functional	**Fixed unit cell**	**Optimized unit cell**
PBE	1.003(1)	1.004(1)
PW91	1.003(1)	1.004(1)
PBEsol	1.001(1)	1.001(1)
PBE-TS	1.003(1)	1.003(1)
PBE-G6	1.004(1)	1.004(1)
PW91-OBS	1.003(1)	1.001(1)
	Bonds containing hydrogen	
	Fixed unit cell	**Optimized unit cell**
PBE	1.021(4)	1.021(4)
PW91	1.021(4)	1.021(4)
PBEsol	1.034(4)	1.036(4)
PBE-TS	1.021(4)	1.022(4)
PBE-G6	1.022(4)	1.023(5)
PW91-OBS	1.022(4)	1.025(5)

Table 3 Mean values (Å), for C=O, C-C, C-C (aromatic), N-N, C-N, C=N, C-H, N-H and O-H bonds from Class I and Class II compounds.

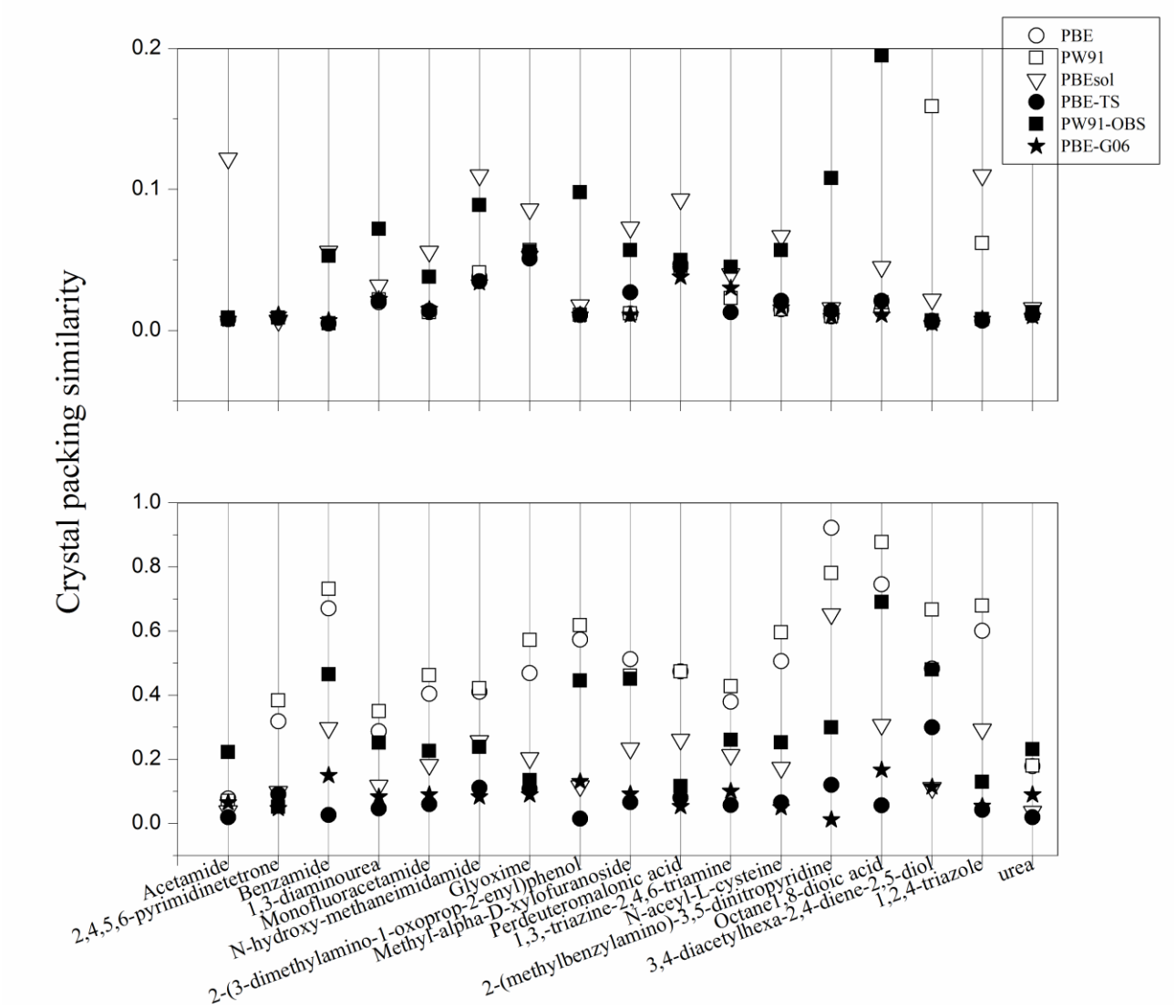
	Mean	
	Fixed unit cell	**Optimised unit cell**
***r*C=O**		
PBE	1.241(21)	1.240(20)
PW91	1.241(21)	1.240(21)
PBEsol	1.242(22)	1.243(22)
PBE-TS	1.241(21)	1.241(21)
PBE-G6	1.241(21)	1.242(21)
PW91-OBS	1.241(21)	1.242(22)
Experimental	1.228(21)	
Number of Observations	20	
***r*C-C**		
PBE	1.496(51)	1.499(52)
PW91	1.496(51)	1.499(52)
PBEsol	1.491(50)	1.492(50)
PBE-TS	1.497(51)	1.496(51)
PBE-G6	1.498(51)	1.496(51)
PW91-OBS	1.496(51)	1.490(50)
Experimental	1.499(50)	
Number of Observations	66	
***r*C-C (Aromatic)**		
PBE	1.402(20)	1.404(20)
PW91	1.401(20)	1.403(20)
PBEsol	1.400(19)	1.401(19)
PBE-TS	1.403(20)	1.402(20)
PBE-G6	1.404(20)	1.403(19)
PW91-OBS	1.402(20)	1.398(19)
Experimental	1.405(20)	
Number of Observations	46	
***r*N-N**		
PBE	1.398(21)	1.402(19)
PW91	1.400(22)	1.404(19)
PBEsol	1.392(22)	1.394(21)
PBE-TS	1.398(21)	1.399(19)
PBE-G6	1.399(21)	1.399(20)
PW91-OBS	1.399(21)	1.396(18)
Experimental	1.389(28)	
Number of Observations	6	
***r*C-N**		
PBE	1.393(56)	1.396(57)
PW91	1.393(57)	1.396(57)
PBEsol	1.387(54)	1.388(55)
PBE-TS	1.393(56)	1.393(56)
PBE-G6	1.394(56)	1.393(57)
PW91-OBS	1.393(57)	1.388(56)
Experimental	1.391(61)	
Number of Observations	34	
***r*C=N**		
PBE	1.346(19)	1.347(19)
PW91	1.346(19)	1.346(19)
PBEsol	1.344(19)	1.344(19)
PBE-TS	1.346(19)	1.345(19)
PBE-G6	1.347(19)	1.346(19)

PW91-OBS	1.348(19)	1.344(19)
Experimental	1.336(21)	
Number of Observations	9	
***r*C-H**		
PBE	1.095(5)	1.096(5)
PW91	1.094(5)	1.095(5)
PBEsol	1.101(5)	1.101(6)
PBE-TS	1.095(5)	1.095(5)
PBE-G6	1.096(5)	1.096(6)
PW91-OBS	1.095(5)	1.093(5)
Experimental	1.088(9)	
Number of Observations	140	
***r*N-H**		
PBE	1.034(8)	1.033(8)
PW91	1.034(8)	1.033(8)
PBEsol	1.040(10)	1.042(11)
PBE-TS	1.034(8)	1.034(9)
PBE-G6	1.035(8)	1.036(9)
PW91-OBS	1.035(8)	1.038(10)
Experimental	1.018(9)	
Number of Observations	29	
***r*O-H**		
PBE	1.026(21)	1.023(19)
PW91	1.026(21)	1.024(19)
PBEsol	1.047(27)	1.052(28)
PBE-TS	1.026(22)	1.028(21)
PBE-G6	1.025(21)	1.029(21)
PW91-OBS	1.027(19)	1.036(20)
Experimental	1.002(18)	
Number of Observations	13	

Table 4 The gradient of the best-fit line, *m*, for hydrogen bond length contacts observed for Class I compounds for each DFT functional.

	Fixed unit cell	Optimised unit cell
***r*O...H**		
PBE	0.982(3)	1.043(22)
PW91	0.983(4)	1.023(22)
PBEsol	0.960(7)	0.963(16)
PBE-TS	0.982(4)	0.980(7)
PBE-G6	0.983(3)	0.966(7)
PW91-OBS	0.971(5)	0.959(19)
Number of Observations	27	
***r*O…D**		
PBE	1.000(6)	1.014(8)
PW91	0.993(2)	1.018(13)
PBEsol	0.981(4)	0.982(9)
PBE-TS	0.993(2)	0.990(3)
PBE-G6	0.994(2)	0.982(3)
PW91-OBS	0.984(4)	0.964(6)
Number of Observations	27	
***r*N…H**		
PBE	0.980(14)	1.002(14)
PW91	0.981(14)	1.001(15)
PBEsol	0.964(18)	0.948(21)
PBE-TS	0.987(13)	0.971(15)
PBE-G6	0.983(13)	0.955(16)
PW91-OBS	0.968(15)	0.911(17)
Number of Observations	12	
***r*N…D**		
PBE	0.986(6)	0.992(4)
PW91	0.983(6)	0.989(3)
PBEsol	0.974(9)	0.958(5)
PBE-TS	0.988(6)	0.978(5)
PBE-G6	0.986(6)	0.967(3)
PW91-OBS	0.981(6)	0.951(4)
Number of Observations	12	

(a)


PBE
PW91
PBEsol
PBE-TS
PW91-OBS
PBE-G06
Crystal packing similarity
0.2
0.1
0.0
1.0
0.8
0.6
0.4
0.2
0.0
Acetamide
2,4,5,6-pyrimidinetetrone
Benzamide
1,3-diaminourea
Monofluoracetamide
N-hydroxy-methaneimidamide
Glyoxime
2-(3-dimethylamino-1-oxoprop-2-enyl)phenol
Methyl-alpha-D-xylofuranoside
Perdeuteromalonic acid
1,3,-triazine-2,4,6-triamine
N-aceyl-L-cysteine
2-(methylbenzylamino)-3,5-dinitropyridine
Octane1,8-dioic acid
3,4-diacetylhexa-2,4-diene-2,5-diol
1,2,4-triazole
urea

(b)

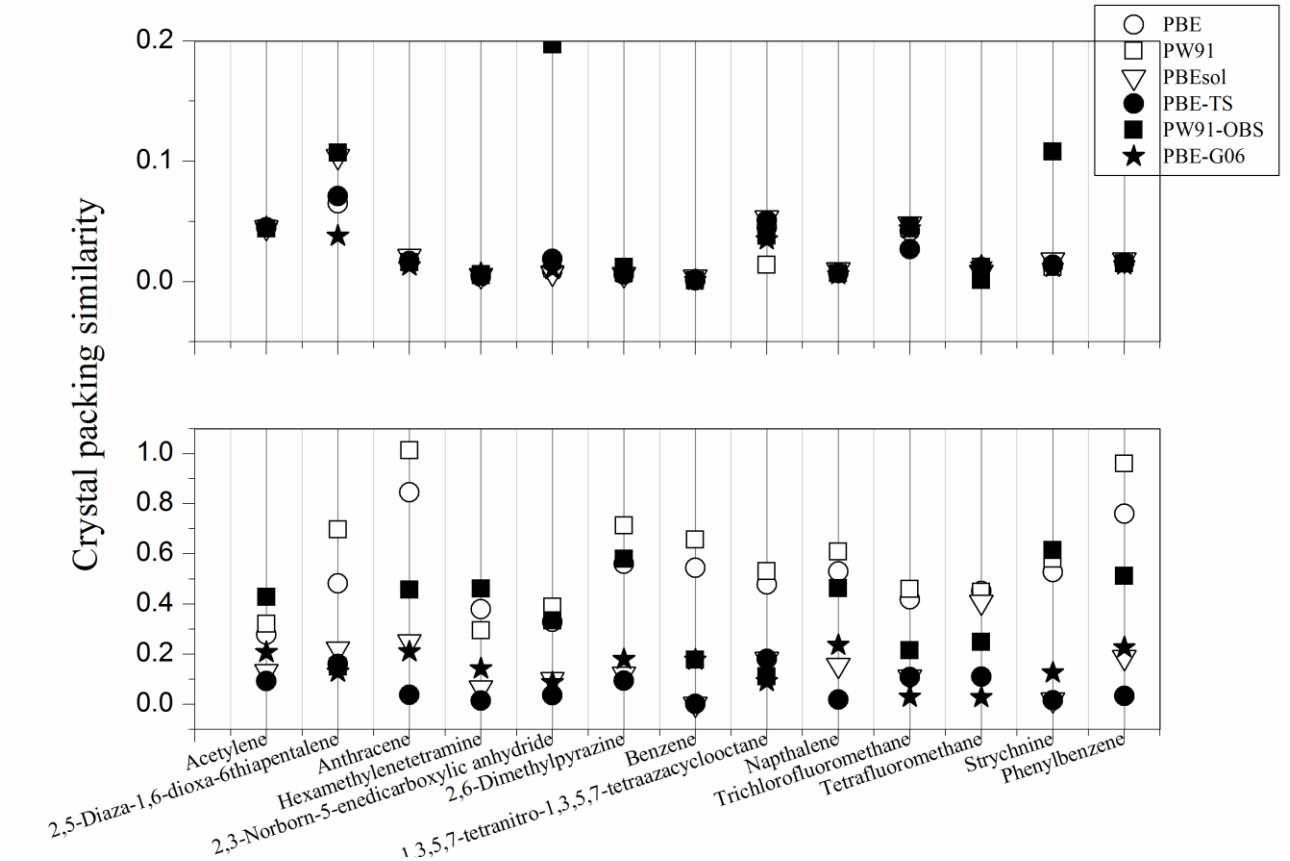
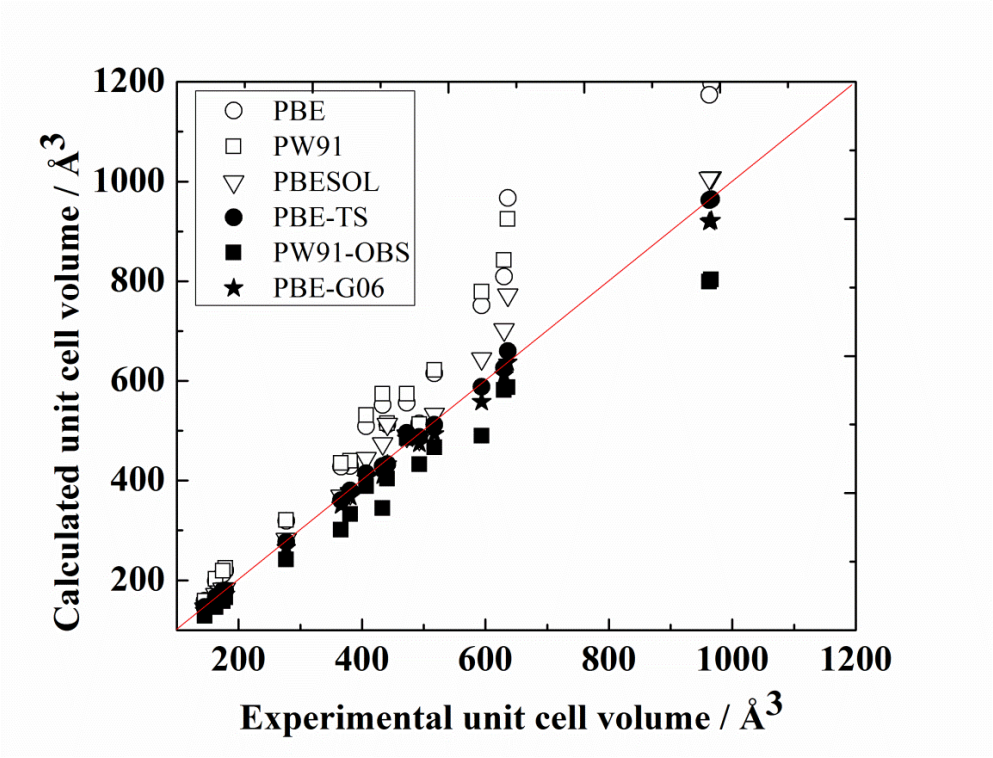


Figure 1 Crystal packing similarity plots for (a) Class I and (b) Class II structures. The top panels refer to data obtained in the fixed cell minimisations, the lower panels to results obtained where the unit cell parameters, in addition to the atom positions, are allowed to optimise. Note the difference in vertical scales in the upper and lower figures.

(a)



(b)

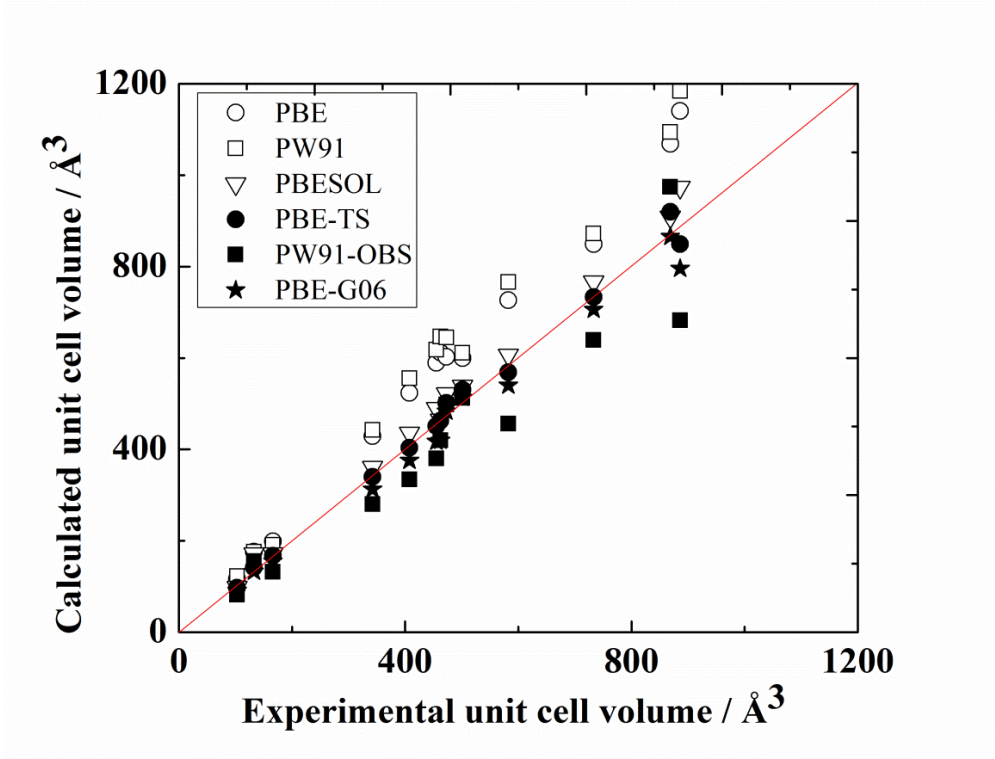
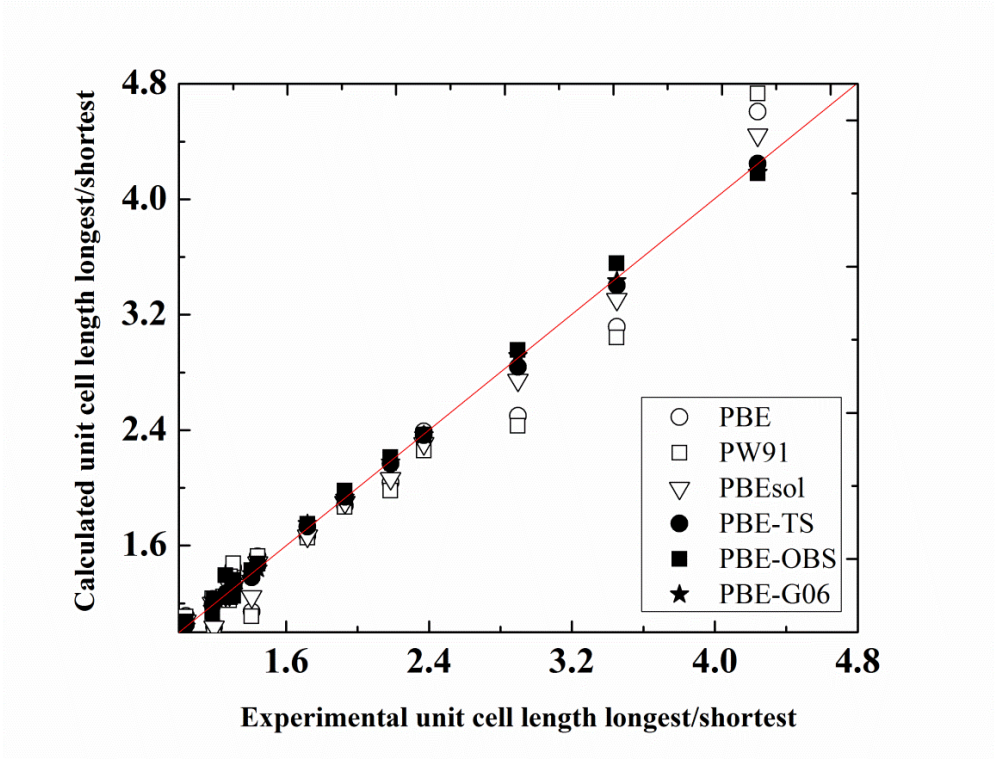


Figure 2 Experimental versus calculated unit cell volumes for (a) Class I and (b) Class II structures. The $y = x$ line is drawn as a visual guide to assess the quality of the simulation results.

(a)



(b)

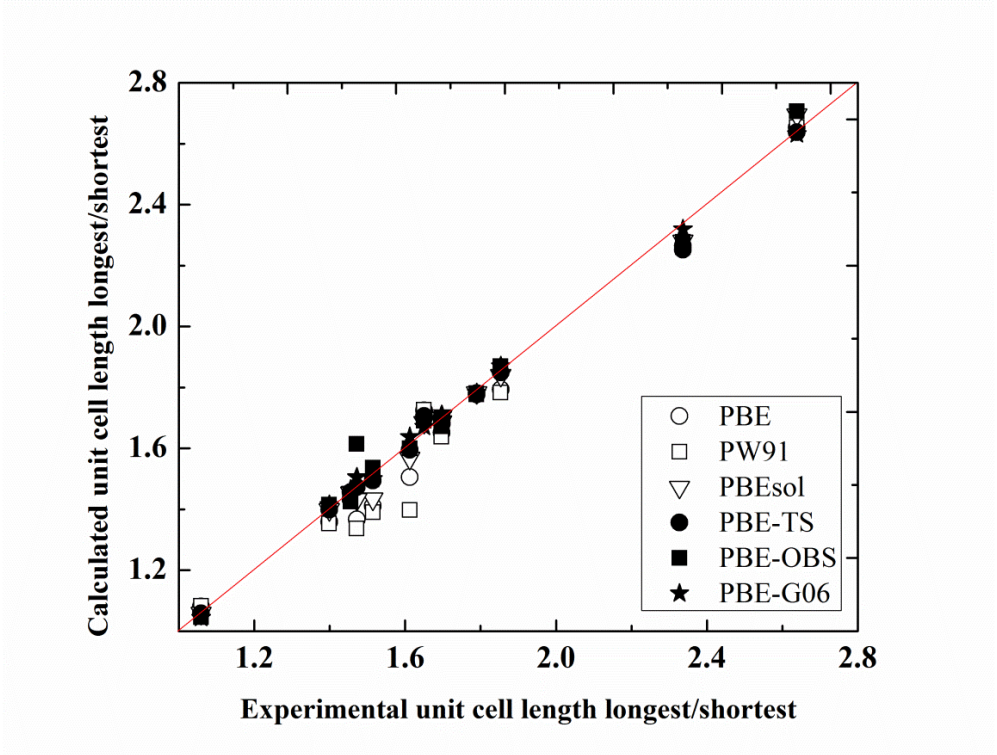


Figure 3 Experimental unit cell length (longest/shortest) vs calculated for (a) Class I and (b) Class II structures. The *y* = *x* line is drawn as a visual guide to assess the quality of the simulation results.